

Einführung in die empirische Wirtschaftsforschung

Prüfung HS 2019

Anweisungen:

- Die Prüfung besteht aus fünfzehn Multiple-Choice-Aufgaben.
- Jede Multiple-Choice-Aufgabe hat genau eine richtige Lösung. Wenn Sie diese ankreuzen, erhalten Sie einen Punkt. Wenn Sie eine falsche Lösung ankreuzen, werden Ihnen 0.33 Punkte abgezogen. Wenn Sie mehr als eine Lösung ankreuzen oder es nicht 100%ig klar ist, welche Lösung Sie angekreuzt haben, erhalten Sie null Punkte.
- **Wichtig: Unterschreiben Sie das Antwortblatt und befolgen Sie die Instruktionen auf der Rückseite!**

1. Man betrachtet das Regressionsmodell

$$y_i = \beta_1 x_i + u_i$$

Eine Zufallsstichprobe mit $n = 5$ ergibt die folgenden Daten (x_i, y_i) :

$(1, 1.2), (2, 2.9), (3, 3.8), (4, 4.0), (5, 5.3)$. Was ist der Kleinste-Quadrate Schätzer für β_1 ?

- (a) $\hat{\beta}_1 = 0.93$
- (b) $\hat{\beta}_1 = 1.11$
- (c) $\hat{\beta}_1 = 1.01$
- (d) $\hat{\beta}_1 = 0.89$

2. Man fragt sich, ob das durchschnittliche Gehalt von BerufsanfängerInnen verschieden ist für Männer und Frauen. Dazu erhebt man eine Zufallsstichprobe von $n = 200$ BerufsanfängerInnen und notiert die zugehörigen Gehälter. Die Daten ergeben einen Stichprobenmittelwert von 58.2 für Männer und einen Stichprobenmittelwert von 56.3 für Frauen (in der Einheit CHF 1'000). Man betrachtet auch das folgende Regressionsmodell:

$$G_i = \beta_0 + \delta D_i + u_i$$

wobei G das Gehalt ist und D eine Dummy-Variable, die den Wert 1 für Frauen und den Wert 0 für Männer annimmt. Was ist der Kleinste-Quadrate Schätzer für δ in diesem Modell (mit den gleichen Daten)?

- (a) $\hat{\delta} = 56.3$
- (b) $\hat{\delta} = 58.2$
- (c) $\hat{\delta} = -1.9$

(d) $\hat{\delta} = 1.9$

3. Sei $\mathbf{X}$ ein Zufallsvektor mit Erwartungswert und Varianz-Kovarianz-Matrix gegeben wie folgt:

$$E(\mathbf{X}) = \begin{pmatrix} -1 \\ 5 \\ -6 \end{pmatrix} \quad \text{und} \quad Var(\mathbf{X}) = \begin{pmatrix} 9 & 3 & 2 \\ 3 & 4 & -2 \\ 2 & -2 & 3 \end{pmatrix}$$

Dann gilt für den Erwartungswert und die Varianz der Zufallsvariablen Y definiert als $Y = \mathbf{a} + \mathbf{b}'\mathbf{X}$, mit $\mathbf{a} = 1$ und $\mathbf{b} = (-1, 0, -2)'$:

- (a) $E(Y) = 14$ und $Var(Y) = 12$
- (b) $E(Y) = 14$ und $Var(Y) = 29$
- (c) $E(Y) = 12$ und $Var(Y) = 13$
- (d) $E(Y) = 12$ und $Var(Y) = 29$

4. Man betrachtet ein multiples Regressionsmodell

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Sei r_{ii} das i -te Diagonal-Element der Matrix $\mathbf{R} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Dann ist das i -te standardisierte Residuum definiert als

$$Stand.-Resid_i = \frac{\hat{u}_i}{s\sqrt{1 - r_{ii}}}$$

Was ist der Vorteil dieser standardisierten Residuen gegenüber den ‘normalen’ Residuen $\hat{u}_i$ (unter der Annahme von Homoskedastie)?

- (a) Die Varianz der standardisierten Residuen ist konstant (mit unbekannten Wert).
- (b) Die Varianz der standardisierten Residuen ist konstant mit Wert 1.
- (c) Der Varianz der standardisierten Residuen ist konstant mit Wert $\sigma^2 = Var(u_i)$.
- (d) Im Gegensatz zu den $\hat{u}_i$ haben die standardisierten Residuen Erwartungswert Null.

5. Eine Forscherin möchte herausfinden, wie der Ernteertrag einer Pflanze, Y , auf die Menge des verabreichten Düngers, X , reagiert. Sie hat dazu $n = 10$ Parzellen (identischen) Bodens zur Verfügung. Auf der i -ten Parzelle wird sie die Menge x_i an Dünger verabreichen, wobei jeweils $x_i \in [0, 10]$ sein muss (in einer angemessenen Einheit). Die Forscherin geht davon aus, dass es eine linear Beziehung der Art

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

gibt und wählt die Werte $\{x_1, \dots, x_{10}\}$ so aus, um die Varianz von $\hat{\beta}_1$ zu minimieren (unter der Annahme von Homoskedastie). Dann gilt für die Varianz von $\hat{\beta}_0$ unter der Annahme von Homoskedastie mit $\sigma^2 = 50$:

- (a) $Var(\hat{\beta}_0) = 0.20$
- (b) $Var(\hat{\beta}_0) = 0.48$
- (c) $Var(\hat{\beta}_0) = 3.16$
- (d) $Var(\hat{\beta}_0) = 10$

6. Hier ist Software-Output für das Beispiel “We All Scream for Icecream” der Vorlesung:

```
Call: lm(formula = sales ~ temp)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 348.8975     10.7646   32.411  < 2e-16 ***
temp         5.9055       0.8265    7.145 8.93e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40.33 on 28 degrees of freedom
Multiple R-squared:  0.6458, Adjusted R-squared:  0.6332
F-statistic: 51.05 on 1 and 28 DF,  p-value: 8.935e-08

> predict(fit.ice, data.frame(temp = 20, sales = 0),
          interval = "confidence", level = 0.95)

      fit      lwr      upr
467.008 443.694 490.322
```

Weiterhin gilt $t_{28,0.975} = 2.048$. Was ist ein 95% Vorhersage-Intervall für die Beobachtung **sales** in einem bestimmten Monat mit **temp** = 20? (Gehen Sie davon aus, dass alle Annahmen, die für die Gültigkeit eines solchen Intervalles nötig sind, zutreffen.)

- (a) [443.7, 490.3]
- (b) [395.7, 558.3]
- (c) [381.2, 552.6]
- (d) Man benötigt zusätzliche Informationen, um dieses Vorhersage-Intervall zu berechnen.

7. Sei $X^* = 32 + 1.8 \cdot X$ und $Y^* = 0.264 \cdot Y$. Wenn $Cov(X, Y) = 485$, was gilt dann für $Cov(X^*, Y^*)$? (Anwendung: X ist Temperatur in Celsius und X^* die gleiche Temperatur in Fahrenheit; Y ist verkaufte Menge von Eiskrem in Liter und Y^* die gleiche Menge in Gallonen.)

- (a) $Cov(X^*, Y^*) = 230.5$
- (b) $Cov(X^*, Y^*) = 262.5$
- (c) $Cov(X^*, Y^*) = 1'001.0$
- (d) $Cov(X^*, Y^*) = 1'033.0$

8. Betrachten Sie das Regressionsmodell

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + u_i$$

Von Interesse ist die Null-Hypothese $H_0 : 0.5\beta_1 + \beta_2 = 2$. Zugehöriger Software-Output ist wie folgt:

Call: `lm(formula = y ~ x1 + x2)`

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.02381	0.11333	-0.210	0.834
x1	2.02753	0.10530	19.256	< 2e-16 ***
x2	0.96095	0.11424	8.412	3.55e-13 ***

Residual standard error: 1.133 on 97 degrees of freedom
Multiple R-squared: 0.8182, Adjusted R-squared: 0.8145
F-statistic: 218.3 on 2 and 97 DF, p-value: < 2.2e-16

Call: `lm(formula = I(y - 4 * x2) ~ I(x1 - 2 * x2))`

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.02221	0.12264	-0.181	0.857
I(x1 - 2 * x2)	1.63678	0.05398	30.323	<2e-16 ***

Residual standard error: 1.226 on 98 degrees of freedom
Multiple R-squared: 0.9037, Adjusted R-squared: 0.9027
F-statistic: 919.5 on 1 and 98 DF, p-value: < 2.2e-16

Call: `lm(formula = I(y - 4 * x1) ~ I(x2 - 2 * x1))`

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.02387	0.11277	-0.212	0.833
I(x2 - 2 * x1)	0.98171	0.04730	20.755	<2e-16 ***

Residual standard error: 1.128 on 98 degrees of freedom
Multiple R-squared: 0.8147, Adjusted R-squared: 0.8128
F-statistic: 430.8 on 1 and 98 DF, p-value: < 2.2e-16

Der zugehörige Wert der Test-Statistik ist

- (a) $F = 86.12$
- (b) $F = 17.75$
- (c) $F = 1.87$
- (d) $F = 0.14$

9. Prof. Werwolf unterrichtet Statistik II an der Universität von Transsylvanien. Im Kurs gibt es insgesamt 100 Studierende. Alle haben eine Midterm-Prüfung (in der Mitte des Semesters) und eine Final-Prüfung (am Ende des Semesters) geschrieben. Die Daten ergeben die folgenden Kennzahlen. Midterm: Mittelwert = 230 und SA = 25; Final: Mittelwert = 150 und SA = 20. Eine einfache Regression mit Final als Zielvariable (y) und Midterm als Regressor (x) ergibt einen geschätzten Abschnitt von 21.2. Zusätzlich haben Sie die Information, dass $\sum_{i=1}^n x_i^2 = 5'351'875$

Finden Sie ein 95% Vorhersage-Intervall für den Final für eine bestimmte Studentin mit einem Midterm von 250 (Gehen Sie davon aus, dass alle Annahmen, die für die Gültigkeit eines solchen Intervalles nötig sind, zutreffen.)

- (a) [157.6, 164.8]
- (b) [132.8, 189.6]
- (c) [127.6, 194.8]
- (d) Man benötigt zusätzliche Informationen, um dieses Vorhersage-Intervall zu berechnen.

10. Man erwartet eine nicht-lineare Beziehung zwischen Y und X in dem Sinne, dass $E(Y)$ eine zweimal differenzierbare konkave Funktion von X ist (d.h. zweite Ableitung negativ). Welche der folgenden Spezifikationen ist **NICHT** geeignet, eine solche Beziehung zu modellieren?

- (a) $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$
- (b) $y_i = \beta_0 + \beta_1 x_i + \beta_2 \sqrt{x_i} + u_i$
- (c) $y_i = \beta_0 + \beta_1 \log(x_i) + u_i$
- (d) $\log(y_i) = \beta_0 + \beta_1 x_i + u_i$

11. Betrachten Sie das folgende Regressionsmodell

$$\log(y) = \beta_0 + \beta_1 \log(x) + u$$

Welche der folgenden Aussagen trifft zu?

- (a) Steigt x um ein Prozent, dann steigt y um $0.1\beta_1\%$
- (b) Steigt x um ein Prozent, dann steigt y im Mittel um $100\beta_1\%$
- (c) Steigt x um ein Prozent, dann steigt y im Mittel um $\beta_1\%$
- (d) Steigt x um ein Prozent, dann steigt y um $\beta_1\%$

12. Die Trainerlegende Jürgen Flopp hat einen Statistik Kurs belegt und will nun sein Wissen in der Praxis anwenden. Dafür liess er seinen schnellsten Spieler, Mohamed Lama, 1000 Mal einen 100 Meter Sprint laufen. Flopp mass bei jedem der 1000 Sprints die Zeit die Lama für den 100 Meter Sprint brauchte. Mit dieser Zufallsstichprobe bildete Flopp das folgende 95% Konfidenzintervall für den Mittelwert der Sprintdauer

$$[9.41, 10.39]$$

Welche der folgenden Hypothesen H_0 kann auf dem 5% Niveau zugunsten der H_1 verworfen werden?

- (a) $H_0: \mu = 9.5, H_1: \mu \neq 9.5$
- (b) $H_0: \mu = 10.35, H_1: \mu < 10.35$

(c) $H_0: \mu = 9.6, H_1: \mu \neq 9.6$

(d) $H_0: \mu = 9.6, H_1: \mu < 9.6$

13. Betrachten Sie das lineare Regressionsmodell ohne Achsenabschnitt

$$y_i = \beta_1 x_i + u_i.$$

Gehen Sie davon aus das die Annahmen des einfachen linearen Regressionsmodelles erfüllt sind. Anstelle von dem OLS Schätzer betrachten Sie folgenden Schätzer für β_1

$$\hat{\beta}_1 = \frac{\bar{y}}{\bar{x}},$$

wobei $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ und $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Ist dieser Schätzer erwartungstreu?
Hinweis: Berechnen Sie $E(\hat{\beta}_1)$

(a) Der Schätzer wäre erwartungstreu wenn das Modell einen Achsenabschnitt hat.

(b) Nein

(c) Ja

(d) Keine Aussage möglich

14. Welche Aussage bezüglich dem linearen Regressionsmodell ist generell falsch. Gehen Sie bei der beantwortung der Frage davon aus das der Fehlerterm einen Erwartungswert von 0 hat und die Regressoren deterministisch sind.

(a) Das R^2 wird beim Hinzufügen von Regressoren (Variablen) niemals kleiner.

(b) Wenn $\beta_1 = 0$ im einfachen linearen Regressionsmodell, dann gilt $R^2 = 0$.

(c) Die heteroskedastisch robust geschätzten Standardfehler (HC3) sind immer grösser gleich den geschätzten Standardfehlern unter der Annahme von Homoskedastie.

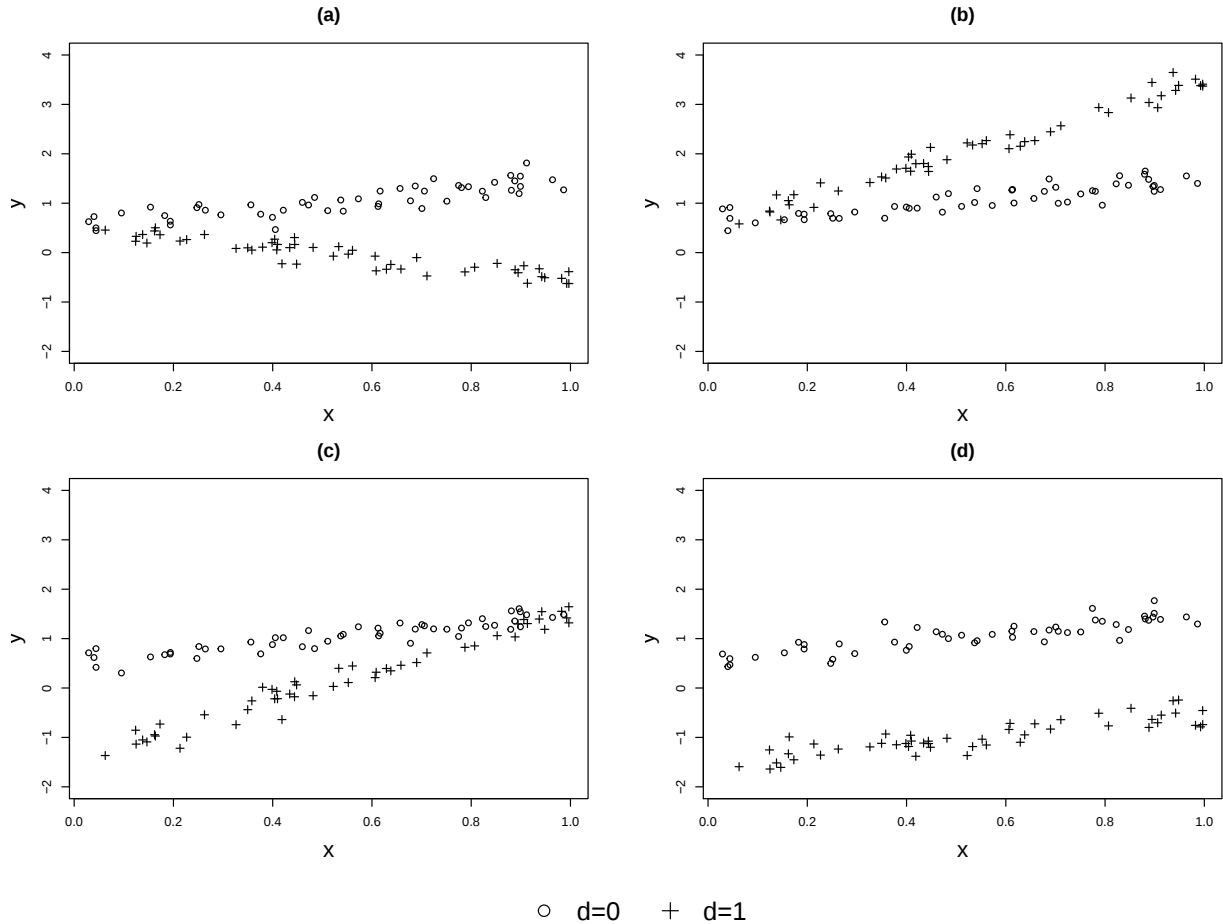
(d) Auch unter Heteroskedastie ist $\hat{\beta}$ ein unverzerrter Schätzer für β .

15. Betrachten Sie das lineare Regressionsmodell

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i d_i + u_i,$$

wobei d_i eine Dummy-Variable ist ($d_i \in \{0, 1\}$), $\beta = (0.5, 1, 2)'$ und $u_i \stackrel{iid}{\sim} N(0, 0.15)$.

Welcher unten dargestellte Datensatz mit $n = 100$ stammt am ehesten vom in der Aufgabe beschriebenen Modell?



Zusätzliche Aufgabe:

2. Man erhebt Daten (x_i, y_i) von einer Zufallsstichprobe von $n = 100$ Autos: X ist das Gewicht gemessen in Kilo und Y ist der (durchschnittliche) Benzinverbrauch gemessen in Litern/100 Kilometer. Es ergibt sich eine Stichprobenkorrelation der (x_i, y_i) von $r = 0.75$. Wenn man *mit den gleichen Daten* den Benzinverbrauch umwandelt in die Einheit Kilometer/Liter (d.h. $Y^* = 1/(100Y)$), was kann man bzgl. der neuen Stichprobenkorrelation r^* der (x_i, y_i^*) erwarten (ohne die individuellen Daten im Detail zu wissen)?

- (a) $r^* > 0$, aber genauer Wert unbekannt
- (b) $r^* = 0.75$
- (c) $r^* < 0$, aber genauer Wert unbekannt
- (d) $r^* = -0.75$